Claims

[c1]   What is claimed is:

1.A method for computerized extracting of scheduling information from a natural language text for automatic entry into a calendar application, the method comprising the following steps:

(a) parsing the natural language text to build a dependency tree;

(b) determining if the natural language text contains scheduling information by calculating a probability sum for the dependency tree; and

(c) if the probability sum exceeds a predetermined value, extracting scheduling information from the dependency tree and exporting the scheduling information to the calendar application.

[c2]   2.The method of claim 1, wherein parsing the natural language text further comprises segmenting each sentence in the natural language text into words,building the dependency tree containing dependency pairs by comparing word pairs in the natural language text with a dependency database, and adding the word pairs found in the dependency database as dependency pairs to the

dependency tree.

[c3]   3.The method of claim 2, wherein when building the dependency tree:

for each sentence in the natural language text, forming a head word list of all possible head words in the sentence that are specified in a master positionover the sentence; and

pairing each word in each sentence in the natural language text with the possible head words in the head word list to form a word pair, wherein if the word pair formed by the word and the possible head word is found in the dependency database, adding the word pair formed by the word and the possible head word as a dependency pair to the dependency tree.

[c4]   4.The method of claim 2, further comprising building the dependency database using the following steps:

segmenting each sentence in a text corpus into words, wherein the text corpus contains a plurality of sample natural language texts containing scheduling information;

for each sentence in the text corpus, checking all possible combinations of word pairs to determine if the word pair has a high co-occurrence in the text corpus;

if the word pair has a high co-occurrence in the text corpus, determining the head word using a tagged corpus,

and checking the validity of the word pair using violation constraints, wherein the tagged corpus specifies the actual head words for sentences relevant to scheduling information in the text corpus and contains dependencies for all other words with respect to the actual head words, and the violation constraints specify illegal dependency structures;

if the word pair is a valid dependency pair, computing the probability of the word pair, adding the word pair as a dependency pair to the dependency database, and adding the probability of the dependency pair to the dependency database, wherein the probability of the dependency pair corresponds to the frequency of the word pair in the text corpus; and

repeating the above steps until no new dependency pairs are identified.

[c5] 5.The method of claim 2, wherein determining if the natural language text contains scheduling information further comprises calculating a probability sum for the natural language text by adding up probabilities for all the dependency pairs in the dependency tree, the probability of each dependency pair corresponding to the frequency of the dependency pair in a text corpus, the text corpus containing a plurality of sample natural language texts containing scheduling information.

[c6]   6.The method of claim 1, wherein after extracting scheduling information from the natural language text, the method further comprising computing a value for the scheduling information.

[c7]   7.The method of claim 1, wherein after extracting scheduling information from the natural language text, the method furthercomprising sending a confirmation message to a user to confirm the scheduling information.

[c8]   8.The method of claim 1, wherein exporting the extracted scheduling information to the calendar application further comprises sending a confirmation message to the calendar application.

[c9]   9.The method of claim 1, wherein the natural language text is a natural language e-mail.

[c10]   10.A personal organization apparatus comprising:
a processor for executing code in the personal organization apparatus; and
a storage unit connected to the processor for storing data used by the processor including a natural language text;
wherein the processor parses the natural language text to build a dependency tree in the storage unit, deter-

mines if the natural language text contains scheduling information by calculating a probability sum for the dependency tree, and if the probability sum exceeds a predetermined value, extracts scheduling information from the dependency tree and exports the scheduling information to a calendar application.

[c11]   11. The personal organization apparatus of claim 10, wherein the storage unit further includes a dependency database, the dependency database specifying a plurality of dependency pairs and the corresponding probability of each dependency pair, each dependency pair being a word pair found in a text corpus, the probability of the dependency pair corresponding to the frequency of the word pair in the text corpus, and the text corpus including a plurality of sample natural language texts containing scheduling information; and

wherein the processor builds the dependency tree in the storage unit containing dependency pairs by comparing word pairs in the natural language text with the dependency database and adding the word pairs found in the dependency database as dependency pairs to the dependency tree, calculates a probability sum for the natural language text by adding up probabilities for all the dependency pairs in the dependency tree, and if the probability sum exceeds a predetermined sum, extracts

scheduling information from the dependency tree and exports the scheduling information to a calendar application.

[c12] 12.The personal organization apparatus of claim 11, wherein when the processor builds the dependency tree: for each sentence in the natural language text, the processor forms a head word list of all possible head words in the sentence that are specified in a master position over the sentence, the head word list being stored in the storage unit; and

the processor pairs each word in each sentence in the natural language text with the possible head words in the head word list, wherein if the word pair formed by the word and the possible head word is found in the dependency database, the processor adds the word pair formed by the word and the possible head word as a dependency pair to the dependency tree.

[c13] 13.The personal organization apparatus of claim 11, wherein the processor builds the dependency database using the text corpus, wherein for each sentence in the text corpus, the processor checks all possible combinations of word pairs to determine if the word pair has a high co-occurrence in the text corpus; if the word pair has a high co-occurrence in the text corpus, the processor determines the head word using a tagged corpus,

and checks the validity of the word pair using violation constraints, wherein the tagged corpus specifies actual head words for sentences relevant to scheduling information in the text corpus and contains dependencies for all other words with respect to the actual head words, and the violation constraints specify illegal dependency structures; and if the word pair is a valid dependency pair, the processor determines the frequency of the word pair in the text corpus and adds the word pair as a dependency pair to the dependency database and adds the frequency of the word pair as the probability of the dependency pair to the dependency database.

[c14]   14. The personal organization apparatus of claim 13, wherein the processor repetitively builds the dependency database until no new dependency pairs are identified.

[c15]   15.The personal organization apparatus of claim 13, wherein when building the dependency database the processor further segments each sentence in the text corpus into words.

[c16]   16.The personal organization apparatus of claim 10, wherein the processor further segments each sentence in the natural language text into words.

[c17]   17.The personal organization apparatus of claim 10,

wherein after extracting scheduling information from the natural language text, the processor computes a value for the scheduling information.

[c18]   18. The personal organization apparatus of claim 10, wherein after extracting scheduling information from the natural language text, the processor sends a confirmation message to a user interface module to confirm the scheduling information.

[c19]   19. The personal organization apparatus of claim 10, wherein when the processor exports the scheduling information to the calendar application, the processor further sends a confirmation message to the calendar application.

[c20]   20. The personal organization apparatus of claim 10, wherein the natural language text is a natural language e-mail.